

AMENDMENTS TO THE CLAIMS:

This listing of claims will replace all prior versions, and listings, of claims in the application. Please cancel claims 1-20 without prejudice.

LISTING OF CLAIMS:

Claims 1-20 (Cancelled)

21. (Previously Presented) A method of storing a natural language text corpus in a database, comprising the steps of:

identifying word tokens of said natural language text corpus;

determining locations in the natural language text of the identified word tokens;

determining word types associated with the identified word tokens;

storing the determined word types in said database, wherein the number of stored word types is less than the number of identified word tokens;

storing word token location identifiers identifying the determined locations in the natural language text corpus of the identified word tokens; and

linking the stored word token location identifiers to the stored word types, such that, for a given identified word token, the stored word token location identifier identifying the location of the identified word token is logically linked to the stored word type associated with the identified word token.

22. (Previously Presented) The method according to claim 21, further comprising the steps of:

determining morpho-syntactic descriptions for the identified word tokens;

storing the morpho-syntactic descriptions for the identified word tokens; and

linking the stored morpho-syntactic descriptions to the stored word token location identifiers, such that, for a given identified word token, the stored morpho-syntactic description for the identified word token is logically linked to the stored word token location identifier identifying the location of the identified word token.

23. (Previously Presented) The method according to claim 22, wherein the morpho-syntactic description of a word token comprises a part-of-speech and an inflectional form of the word token.

24. (Previously Presented) The method according to claim 21, further comprising the steps of:

identifying phrases of said natural language text corpus;

determining word tokens comprised in the identified phrases; and

storing phrase location identifiers identifying the stored word token location identifiers of the word tokens comprised in the identified phrases, such that, for a given identified phrase, the stored phrase location identifier of the identified phrase identifies the stored word token location identifiers identifying the location of the word tokens comprised in the identified phrase.

25. (Previously Presented) The method according to claim 24, further comprising the steps of:

determining phrase types of the identified phrases;

storing the determined phrase types; and

linking the stored phrase types to the stored phrase location identifiers, such that, for a given identified phrase, the phrase type for the identified phrase is logically linked to the stored phrase location identifier identifying the stored word token location identifiers identifying the location of the word tokens comprised in the identified phrase.

26. (Previously Presented) A system for storing a natural language text corpus, comprising:

a text analysis unit for identifying word tokens of said natural language text corpus, determining locations in the natural language text of the identified word tokens, and determining word types associated with the identified word tokens;

a database for storing the determined word types, wherein the number of stored word types is less than the number of identified word tokens, storing word token location identifiers identifying the location in the natural language text corpus of a respective identified word token, and linking the stored word token location identifiers to the stored word types, such that, for a given identified word token, the stored word token location identifier identifying the location of the identified word token is logically linked to the stored word type which is associated with the identified word token.

27. (Previously Presented) The system according to claim 26, wherein the text analysis unit is further adapted to determine morpho-syntactic descriptions for the identified word tokens, and the database further stores the morpho-syntactic descriptions for the identified word tokens, and links the morpho-syntactic descriptions to the stored word type location identifiers, such that, for a given identified word token, the morpho-syntactic description for the identified word token is logically linked to the stored word token location identifier identifying the location of the identified word token.

28. (Previously Presented) The system according to claim 27, wherein the morpho-syntactic description for the word token comprises a part-of-speech and an inflectional form of the word token.